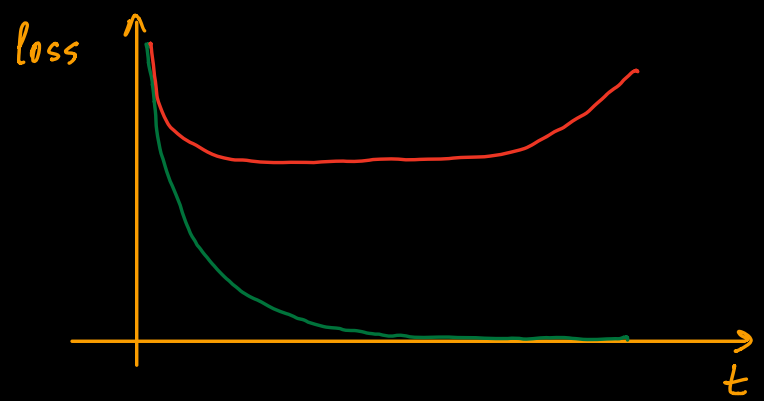
$x \sim X$ | SGD |

1). $b \sim 2^X$, $d \sim B(p)$.

2) $W = W - \lambda \cdot \nabla_W L(W, b)$.

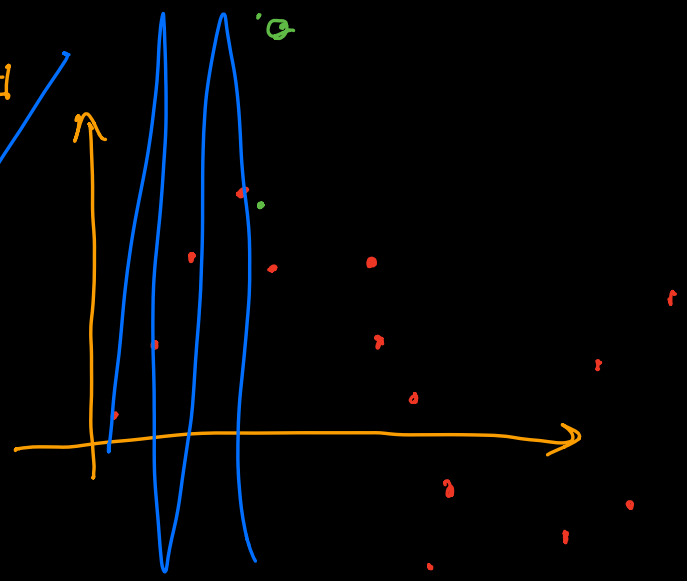


val.

train.

$\tilde{L} = L + P \cdot R$

$R = \|W\|$



$$h_1 = \psi(W_1^T \underbrace{x}_{n,512}) \cdot d_1 \cdot \frac{1}{p} \qquad h_i \in R^{512}$$

$$h_2 = \psi(W_2^T \underbrace{h_1}_{512 \times 512}) \cdot d_2 \cdot \frac{1}{p} \qquad \hat{h}_i = h_i \cdot d \cdot \frac{1}{p}$$

$$\vdots$$

$$h_n = \psi(W_n^T h_n) \cdot d_n \cdot \frac{1}{p}$$

$d \sim B(p)^{512}$

$d \in R^{512}$

$$\begin{cases} 1, & p \\ 0, & 1-p \end{cases}$$